REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 12-01-2004

2. REPORT TYPE Final Report

3. DATES COVERED *(From – To)* 10 September 2002 - 10-Sep-03

4. TITLE AND SUBTITLE

The Combined Roles of Low-Level Perception and Expectation in Conceptual Learning

5a. CONTRACT NUMBER FA8655-02-M4028

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Fernand Gobet

5d. PROJECT NUMBER

5d. TASK NUMBER

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Nottingham
School of Psychology
Nottingham NG7 2RD
United Kingdom

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
PSC 802 BOX 14
FPO 09499-0014

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S) SPC 02-4028

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report results from a contract tasking University of Nottingham as follows: The contractor will investigate the interplay between low-level feature extraction and high-level concept formation, using "chunking mechanisms." The model will be presented with pictures of various objects, and will extract the relevant features before using them to build up a discrimination network for classifying novel pictures. The high-level chunks acquired by the model will be used to direct eye movements, which in turn will determine what will be learned. Questions of interest will include how robust the model is to current expectations, noisy input (to simulate poor visibility) and varying object orientation. The behavior of the model will be compared to human behavior, e.g., learning to categorize images of planes, object identification from aerial photographs, etc. A successfully implemented model will help predict factors which affect human performance in visual recognition, and suggest support techniques for anticipating and correcting errors.

15. SUBJECT TERMS

EOARD, Cognition

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18, NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLAS	UNCLAS	UNCLAS	UL	12	VALERIE E. MARTINDALE, Lt Col, USAF
					19b. TELEPHONE NUMBER *(Include area code)* +44 (0)20 7514 4437

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39-18

Final Report on EOARD Award: FA8655-02-M-4028

Combining Low-Level Perception and Expectations in Conceptual Learning

Fernand Gobet and **Peter Lane**

Introduction

Objectives

The main objective of this one-year project was to extend an established model of human perception, CHREST (Gobet et al, 2001), so that it could combine what it was seeing with what it was expecting to see. This objective has been successfully met, and experiments in a digit-recognition task demonstrate some key phenomena in how humans combine expectations and visual information. The critical technical enhancement made to the CHREST model was a capability to handle both visual and verbal sources of information, and make links between the two.

Our model fills a gap (as observed by Ritter et al, 2003) in previous models of human perception, which fail to capture how low-level visual recognition is guided by semantically-driven expectations. On-going work will expand the range of possible semantic relations, and also the power of CHREST's ability to recognise low-level visual objects. Future applications in modelling human behaviour and image analysis will take advantage of the mechanisms for learning cross-modal links between perceptual and conceptual information developed within this project.

Personnel

The award supported a Research Assistant, Mr. Anthony Sykes, who was employed for one year on a 60% FTE basis. Dr. Fernand Gobet and Dr. Peter Lane managed the project and contributed to its development and dissemination.

Main Results

The technical proposal had two main objectives: first to use CHREST to learn cross-modal relations between visual and verbal forms of input, and second to model human perception under difficult conditions. We discuss the results achieved for each of these two objectives.

Learning Cross-Modal Relations with CHREST

The main result of this project was the extension of CHREST to use multiple forms of input data, in particular, visual and verbal information. In order to develop a model which can interpret and report on visual information in a meaningful manner, it is first important that the model be capable of learning and using links between the different forms of input. The manner in which this is done is summarised in Figure 1.

As illustrated, the CHREST model is divided into three major components. The first handles the input to the model. Previous work had concentrated on input from a simulated eye, i.e. visual

information, only. We extended the input to allow simultaneous input from a simulated verbal channel. In this case, the verbal information *names* the object presented to the eye. The second major component is the pair of short-term memories (STMs). Because CHREST now has two different modalities of input, we had to include a new STM for the verbal information. The final component in CHREST is the long-term memory (LTM). The LTM stores familiar patterns as nodes within a self-organising discrimination network.

In order to use information from the multiple input modalities, CHREST had to be extended to associate and combine these different kinds of input data. We achieved this with *naming links*. A naming link is a link between a familiar visual pattern and a familiar verbal pattern. These links provide the mechanism by which information is transferred between visual and verbal representations of perceived objects.

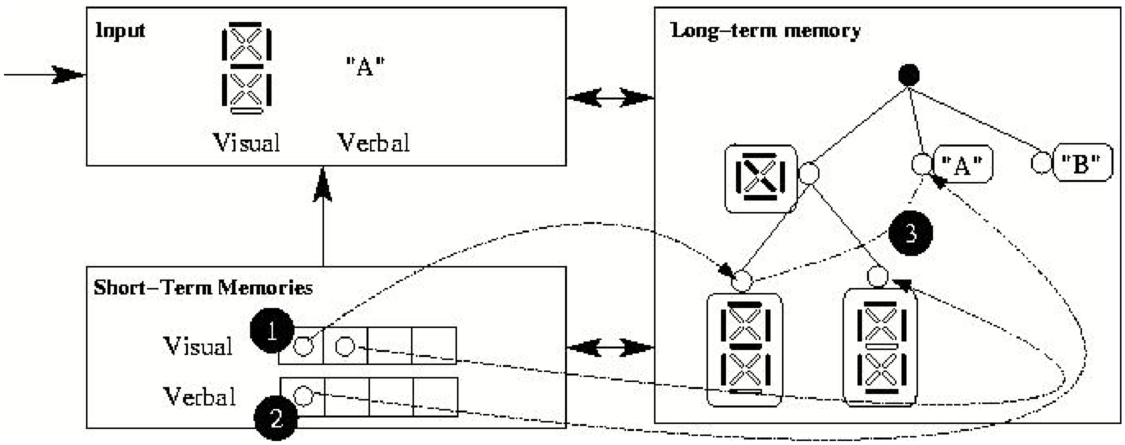


Figure 1 : Learning to link information across two modalities. (1) The visual pattern is sorted through LTM, and a pointer to the node retrieved placed into visual STM. (2) The verbal pattern is sorted through LTM, and a pointer to the node retrieved placed into verbal STM. (3) A naming link is formed between the two nodes at the top of the STMs. (After Lane, Sykes & Gobet, 2003.)

Modelling Human Perception under Difficult Conditions

Using a digit-recognition task, computer simulations with our model have demonstrated three key phenomena observed in human perception:

1. classification accuracy is improved for expected objects;
2. recognition of familiar objects is faster than for unfamiliar ones; and
3. interpreting a set of ambiguous objects is possible with the use of a schema from a different input modality.

We now reference some of the background literature emphasising the importance of these phenomena, and give the empirical results of our studies with the CHREST model.

Background Literature

We focus on three important phenomena demonstrating the role of expectations in perception. The first of these is that expected objects are recognised with greater accuracy than unexpected objects, particularly in domains where noise affects the quality of the input stimulus. For instance, characters may be badly formed, ambiguous, or simply ‘damaged’ or partially hidden. An

expectation that characters are from a standard alphabet enables correct identification of characters which would otherwise be ambiguous (Neisser, 1966; Richman & Simon, 1989).

The second phenomenon shows that expectations may relate to complex collections of objects, or *schemata*. Perceptual classification of objects within a familiar schema can be quicker than when the objects are not in the schema. For instance, Biederman (1981) describes an experiment in which participants took longer to identify a fire-hydrant when positioned above street level in an image than when at its expected position. A similar result can be found in reading: identifying the 'K' in a word such as 'ANKLE' is quicker than in a non-word such as 'XGKAL'.

A third phenomenon is that of *reconstructive memory*, whereby a set of partially obscured objects may be identified based on their being recognised as a composite. For instance, a collection of partially obscured characters may be identified as a word (Lindsay & Norman, 1972), even though each individual character may be ambiguous.

Although it must be admitted that some of these phenomena are difficult to cleanly replicate in experimental settings, it is clear that people do not simply scan an input stimulus in a serial fashion whilst looking for a given item. Instead, humans employ higher-order constraints, based on expected schemata, to constrain the range of potential matches. In other words, perception is not a one-dimensional, bottom-up process, as proposed by Marr (1982), but instead interprets what is being seen in the light of what is currently expected.

Using Cross-Modal Links to Explain these Phenomena

We explore how these phenomena occur in a recognition task, where the model is trained to recognise characters and/or words. The words form *schemata*, or expected sequences of characters. The basic mechanism by which expectations affect perception is through 'priming', in which expected characters can be anticipated and recognised faster. We explain the role of priming in sequence recognition by first explaining the classification, or naming, of individual characters, and then how perceptual characters can be primed by a verbal cue.

Naming

Cross-modal links can be used to name an input visual pattern. The process is applied after sorting the visual pattern through LTM. If the node retrieved has a naming link, then the associated verbal pattern is output by the model: the model thus 'names' the input visual pattern. Using this mechanism, it is possible to train CHREST on a succession of characters, and then request the model to name a succession of new characters; the model's success rate is its classification accuracy.

Priming

The model can be 'primed' to recognise a given visual pattern by presenting its name on the verbal input. Sorting the verbal pattern through LTM, CHREST will locate a node and place this node into its verbal STM. If this node has a naming link, then the linked node is used to prime the model. The priming mechanism uses this linked node as follows. When a visual pattern is presented, it is first compared to the image in the linked node. If it matches to within a given tolerance, then the primed node is returned as the matching node. This priming process can allow the model to successfully match input patterns even though, due to noise or unfamiliar features, they would not have been retrieved during the usual sorting process.

Expecting Sequences

In a similar way to priming with a verbal input, sequence links can be used to prime the model to expect a pattern of the same input modality. Thus, if a node for the visual pattern 'T' is retrieved, and it is linked with a sequence link to a node for the visual pattern 'H', then the model will expect the character 'H' to appear next on the input. The priming mechanism described above applies equally to nodes linked through sequence links. In this manner, the model can use a previously learnt schema in either modality to prime its search for further characters.

Empirical Results

We consider three sets of simulations performed with the model. First, we explore the accuracy with which CHREST can classify characters with increasing amounts of noise. Second, we consider the speed with which characters are visually recognised, comparing the speed of 'pure' bottom-up recognition with that of top-down, expectation-driven recognition. Third, we consider how the use of two modalities enables CHREST to disentangle very noisy data when attempting to satisfy high-level constraints. In this report, we present the empirical results of the experiments; for further details on the experiments themselves, the reader is referred to Lane, Sykes and Gobet (2003).

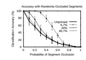
The empirical data used in these experiments is illustrated in the input section of Figure 1: the visual data is a 15-segment array, arranged so that letters can be formed by setting the appropriate segments on and off. Each segment may also be in an indeterminate state, to indicate that it is occluded. The verbal input is a simple character. Sequences of input on either the visual or the verbal inputs are used to represent words.

1. Accuracy of Classification

Our classification task requires CHREST to name the characters appearing in words, with varying amounts of noise added to the perceptual representation. We initially perform the experiment without any priming, so the classification accuracy measures CHREST's ability to recognise the perceptual stimulus, alone. We then ask CHREST to use priming from the verbal input to improve its probability of recognising a noisy or occluded visual input. The tolerance to the match is based on how many segments are permitted to fail to match: results are given for 1 segment (6.7%), 3 (20%) and 7 (46.7%) out of the 15.

Figure 2 shows two graphs, illustrating the classification accuracy of the model when the digits are subjected to two kinds of noise. Figure 2(a) shows the performance when segments are randomly occluded; Figure 2(b) the performance when segments are randomly reversed. The y-axis for both graphs gives the classification accuracy of the model. The x-axis gives the probability of each segment being segment to noise. The experiment was run with probabilities varying from 0.0 to 1.0 in steps of 0.1.

(a)



(b)

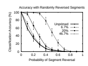


Figure 2 : Classification accuracy results

Both the graphs show similar results: priming improves the model's ability to classify the noisy digits. The lower line is the unprimed performance of the model. The classification accuracy for unprimed input decreases with added noise because some segment states are altered or hidden, and these states will affect the node reached when sorting the visual input.

2. Speed of Classification

We next explored the speed with which classification occurs when characters appear in isolation (unprimed) or within familiar words (primed). We define the *speed* of classification by counting the number of pattern matches made by the model when searching its LTM. For input, we use a set of ten 5-character words. The model is trained until it has learnt the words completely.

Time to recognise characters First, we compare the number of pattern matches required by the model to recognise each of a sequence of characters. 50 models were trained, each with the word-list sorted in a different random order. Table 1 shows the average time, μ, and standard deviation, σ, required when attempting to recognise the characters in isolation (unprimed) as opposed to recognising them when forming part of a word (primed). There is a significant reduction in the required searching time when the model is primed. The use of an expected schema to predict the characters appearing in the visual input significantly reduces classification time.

	μ	σ
unprimed	2.8	1.0
primed	1.4	0.8

Table 1 : Average number of pattern matches

Time to find a given character Second, we consider the time required to find a given character within a sequence of characters. The time required depends on the position of the character within the word, as the previous characters must first be searched. Figure 3 shows the amount of time required to identify a given character in each position: priming means that CHREST identifies and uses a word schema to assist in finding the character, the unprimed timings measure when CHREST treats each character distinctly from its neighbours. CHREST is quicker to locate a character when using a schema, and this advantage increases with character position.

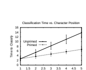


Figure 3 : Relative classification time by position

3. Reconstructive Memory

If the visual input given to the model is such that every character is ambiguous, the only way to attempt a classification is to consider potential schemata which match every character in the scene. For instance, Figure 4 shows three characters, each badly occluded: the first character could be ‘A’ or ‘H’, the second ‘R’ or ‘K’, and the third ‘E’ or ‘F’. From prior familiarity with likely sequences of characters, a viewer may be expected to retrieve the word ‘ARE’ as the likely interpretation, and so assign the appropriate labels to each ambiguous character.

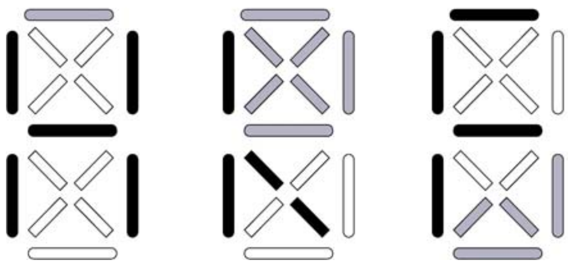


Figure 4 : Reconstructive memory example

We explore CHREST’s ability to reconstruct scenes by training the model on the 26 standard characters, in isolation. Next, CHREST is trained to recognise words using *the verbal input alone*. Thus, CHREST’s LTM has no sequence links between nodes representing visual information. We test CHREST’s ability to reconstruct an interpretation for badly damaged characters by presenting it with sets of words with added noise as visual inputs.

The matching process begins by priming the visual matching process in turn with the named nodes from the LTM’s verbal knowledge. When a match is made to the initial character, the sequence links between the *verbal* schemata are used to further prime the model to match the succeeding characters. Hence, CHREST relies on its familiar schemata to match the noisy data. Figure 5 plots CHREST’s average performance on increasingly noisy visual inputs: three cases are plotted, without using the verbal schemata, with a 20% tolerance and a 46% tolerance. As is evident, the use of priming from the verbal schemata significantly increases the accuracy of CHREST’s performance.

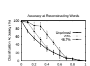


Figure 5 : Reconstruction accuracy

Developed CHREST System

During this project, a graphical interface to the model was developed to facilitate the training, testing and demonstration of the working model. This interface was developed in Java, so as to be portable across all major computer platforms, and is illustrated in Figure 6.

Dissemination

The key theoretical developments were presented at the European Conference of Cognitive Science, held at Osnabruck, Germany from 10th-13th September, 2003. The paper was published in the proceedings (see Lane, Sykes & Gobet, 2003). Attendance at this conference was funded

by the Award. The CHREST model itself was also presented at a tutorial preceding the conference, which was attended by 13 participants. We intend to present similar tutorials at future cognitive-science conferences.

Peter Lane was also awarded a 'Window on Science' grant to visit the US and present the work to the Airforce Academy, in Colorado, and Dr. Kevin Gluck of the Air Force Research Laboratory, in Mesa, Arizona; this visit occurred in August, 2003.

The project has gained some further momentum, and generated two further publications. First, a short article in the AISB Quarterly (Lane & Gobet, 2003). Second, the ideas developed here have contributed towards the larger theory of CHREST as a model for the human mind: these ideas have been presented in a book chapter (Gobet & Lane, in press).

Ongoing Work

The work achieved in this project will be summarised in a journal article, which will be submitted in a few months' time.

The project is continuing with a second EOARD award, no. FA8655-03-1-3071, which will further develop the semantic abilities of the CHREST model of vision. In particular, the project will develop the range of semantic relationships which the model can use, and also how these relations are integrated across the visual and verbal inputs.

References

Publications arising from project

P. C. R. Lane, A. K. Sykes, and F. Gobet, 'Combining Low-Level Perception with Expectations in CHREST', in *Proceedings of EuroCogSci*, pp.205-10, 2003. [See Attached.]

P. C. R. Lane and F. Gobet, 'Towards a model of expectation-driven perception', *AISB Quarterly*, no.114, p.7, 2003.

F. Gobet and P. C. R. Lane, 'The CHREST Architecture of Cognition: Listening to Empirical Data'. In D. Davis (Ed.), *Visions of Mind – Architectures for Cognition and Affect*, to appear.

Other references

I. Biederman, 'On the semantics of a glance at a scene'. In Kubovy, M. & Pomerantz, J. R., editors, *Perceptual Organization*, pages 213–254. Hillsdale, NJ: Lawrence Erlbaum, 1981.

F. Gobet, P.C.R. Lane, S. Croker, P.C-H. Cheng, G. Jones, I. Oliver, and J.M. Pine, 'Chunking mechanisms in human learning', *Trends in Cognitive Sciences*, 5:236-243, 2001.

P. Lindsay & D. Norman, *Human Information Processing*. New York: Academic Press, 1972.

D. Marr, *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman, 1982.

U. Neisser, *Cognitive Psychology*. New York: Appleton-Century-Crofts, 1966.

H. B. Richman, & H. A. Simon, 'Context effects in letter perception: Comparison of two theories'. *PsychologicalReview*, 3:417-432, 1989.

F. E. Ritter, N. R. Shadbolt, D. Elliman, R. Young, F. Gobet and G. D. Baxter, *Techniques for modeling human performance in synthetic environments: A supplementary review*. Wright-Patterson Air Force Base, OH: Human Systems Information Analysis Center, 2003.

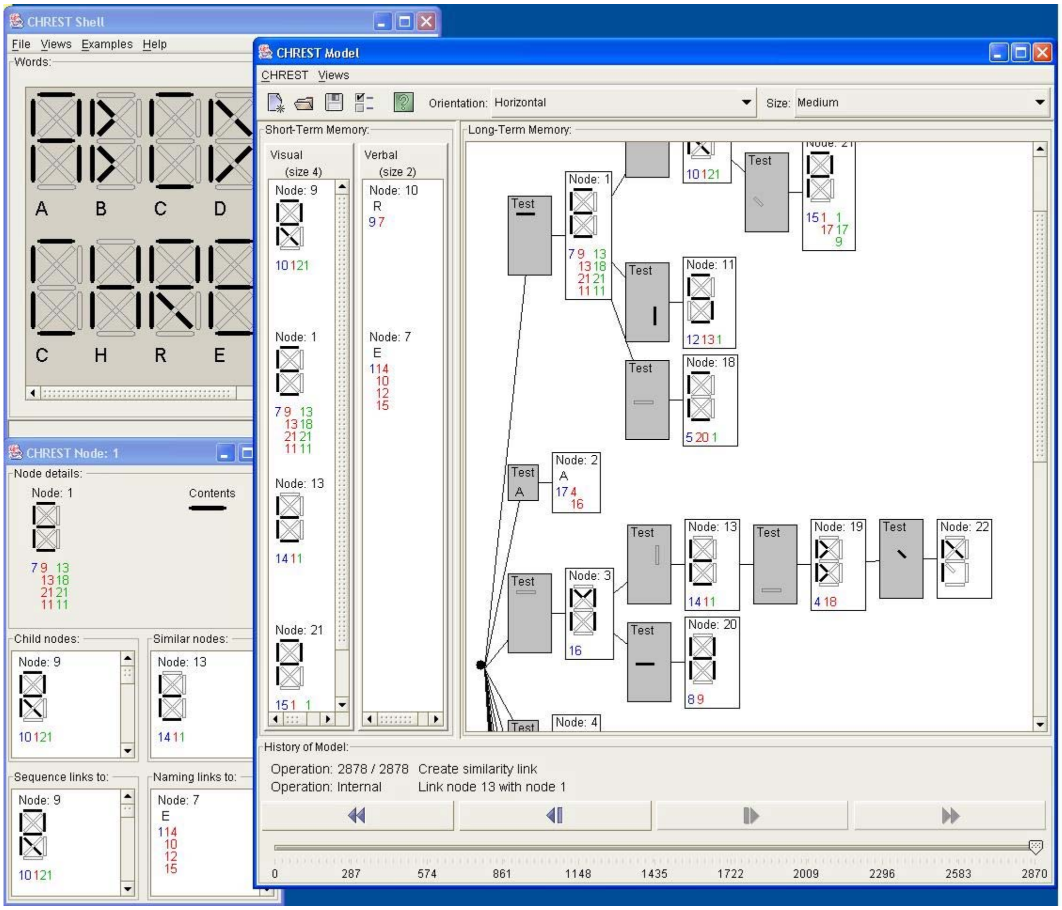


Figure 6 : The Java-based interface to the CHREST model